Kidney International, Vol. 53 (1998), pp. 1446–1448

The relative power of linkage and association studies for the detection of genes involved in hypertension

HYWEL B. JONES

MRC Biostatistics Unit, Cambridge University, Cambridge, England, United Kingdom, and Department of Genetics, Stanford University School of Medicine, Stanford, California, USA

The relative power of linkage and association studies for the detection of genes involved in hypertension. Hypertension is a common disorder that shows a polygenic mode of inheritance. Attempts to localize genes involved in the disorder have been carried out using both linkage and association tests. The relative merit of these two approaches is reviewed with an assessment of their utility for detecting genes involved in hypertension. Power calculations were carried out following the method of Risch and Merikangas [1], assuming markers were typed across the genome. These show that, if there is a single major locus causing susceptibility, non-parametric linkage strategies using affected sibpairs may well prove very effective. However, if there are a number of genes of small effect, the sample size necessary for linkage studies will be prohibitive and a systematic search for allelic association may be more appropriate. This is due to the dramatic reduction in the excess allele sharing for genes of small effect.

Many common disorders show complex patterns of inheritance. This may be due in part to differences in environmental exposure, however, in many cases this complex inheritance will be due to the presence of a number of distinct susceptibility genes. Whether these act independently or in an epistatic fashion will generally be unknown. Because of this, the application of classical linkage analysis that requires the specification of a model of the inheritance is difficult. To circumvent these problems, non-parametric methods of linkage analysis have been developed. Instead of tracking the inheritance of a hypothesized (causative) disease gene, these non-parametric methods examine which parts of the genomes of a pair of affected relatives are identical by descent (IBD) [2]. The most common paradigm uses affected sibling-pairs and compares the IBD allele sharing at a given marker to the expectation under the null hypothesis that no gene is present. Thus, at a marker that is close to a susceptibility gene, it would be expected that there would be an increase in the sharing in that region with affected siblings sharing both alleles IBD more frequently than sharing neither allele.

Risch developed a maximum likelihood method of estimating the IBD sharing probability from sibling-pair data [3–5]. The extent to which sharing deviates from the null depends on both the magnitude of the genetic effect of the susceptibility gene to which the marker is linked (here measured in terms of λ_s) and the strength of that linkage (a function of the recombination fraction, θ). The lod score is then a function of this deviation and the number of sibpairs used in the analysis. The expected deviation from the null for a given θ and λ_s provides an indication of the power to detect susceptibility loci of different effect for maps of different densities. Risch calculated these expected sharing probabilities for a variety of different relative pairs (equations 8 to 22 in [4]). These depend not only on the relative recurrence risk in siblings of affected individuals, but also on the recurrence risk in offspring and monozygotic twins. By assuming that the dominance variance (V_D) is zero (that is, the phenotype of a heterozygote is exactly intermediate to the phenotypes of the two different homozygotes), the expected sharing probabilities may be expressed in terms of λ_s alone [4]. By imposing the condition of no dominance variance, the probability of sharing one allele IBD for a pair of siblings is fixed at 0.5. Consideration of either the deficit in 0-sharing or the excess of 2-sharing will give an indication of the overall deviation from the null hypothesis. The validity of this assumption for hypertension is unclear. The restriction $V_D = 0$ implies that the recurrence risk in siblings is equal to the recurrence risk in offspring of affected parents ($\lambda_s = \lambda_o$) [4]. Because blood pressure is a quantitative age-dependent trait, determining these values (particularly λ_o) is difficult. However, the correlation coefficients have been measured a number of times and have been shown to take similar values for sibling and parent-offspring comparisons ([6] and references therein). As they are of comparable magnitude (0.12 to 0.43 for siblings, 0.12 to 0.37 for parent-offspring), qualitative predictions made under the assumption of no dominance variance will probably be acceptable.

Figure 1 shows the expected 2-sharing (IBD) for a pair of siblings as function of the recombination fraction (θ) between the marker and the susceptibility gene and the magnitude of the genetic effect associated with that particular gene (measured in terms of λ_s). When recombination is zero (the marker is in perfect linkage with the susceptibility gene) and the λ_s is high ($\sim$10), the 2-sharing approaches its maximum possible value (under the assumption of no dominance variance) of 0.5. However, the sharing declines rapidly as the recombination fraction increases, taking the null value of 0.25 at $\theta = 0.5$ irrespective of the magnitude of λ_s.

Whereas the decline in allele sharing as a function of recombination is approximately linear, the reduction in 2-sharing as the λ_s value decreases is initially less extreme. For values greater than 2

Key words: sibling-pair, case control studies, alleles, genome, blood pressure, polygenic inheritance.

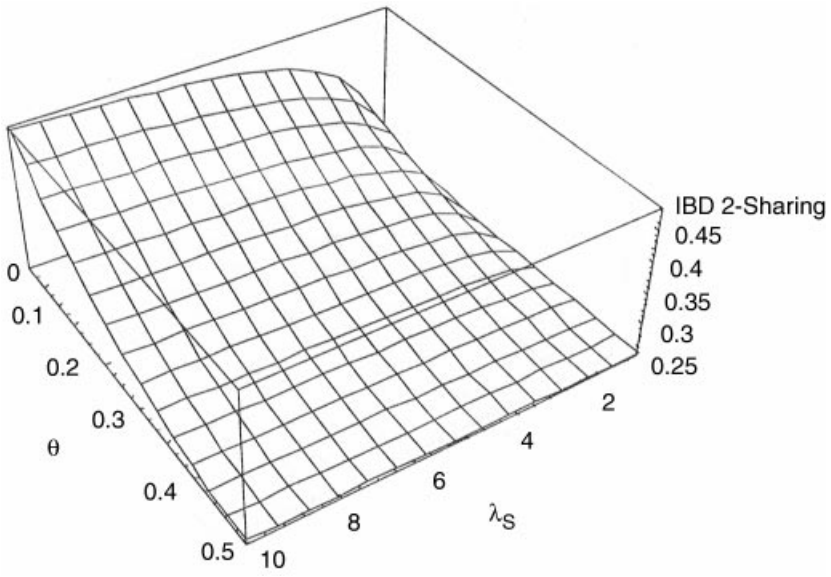

Fig. 1. The expected identical by decent (IBD) 2-sharing for a pair of affected siblings as a function of both the recombination fraction between the disease susceptibility gene and the marker and the λ_s of the susceptibility gene. This was calculated under the assumption of no dominance variance.

the deviation is quite strong, whereas for values less than 2, the sharing rapidly declines towards the null of 0.25. It is important to note that the allele sharing may not always be unambiguously scored, and the power to detect susceptibility genes will be further decreased when there are missing data or low marker polymorphism. Clearly, the density of markers used in the study and their polymorphism is critical to the power of the genome screen [7].

TRANSMISSION DISEQUILIBRIUM TEST

An alternative method of determining the location of a putative disease gene is to look for allelic association. In sibling-pair analysis, the interest is in whether the two affected individuals shared the same allele at a given marker irrespective of which particular allele is actually shared. For association studies, the search is for the over representation of a specific allele in affected individuals. In the simplest scenario, the frequency of alleles in a population of affected individuals (cases) is compared to the frequency in a normal (control) population. A positive result is scored when there is an increased frequency of a particular allele in the cases. The disadvantage of this method is that false positive results can occur when the control population is not appropriately matched to the cases. An alternative test was suggest by Spielman et al [8], which used family based controls. This test requires the collection of trios of two parents and an affected child. The frequency at which alleles are transmitted and not-transmitted to the affected offspring is compared to the Mendelian expectation of 50:50. Importantly, this is a test for linkage disequilibrium requiring the presence of both linkage and allelic association in order to yield a significant result. By using the non-transmitted alleles as the control population, problems of population admixture and mismatched controls are avoided.

STATISTICAL POWER

When designing a genetic study, two factors must be considered: the acceptable false positive (type I) error rate and the false negative (type II) error rate. False positives occur when a result is considered to be statistically significant evidence for the presence of a susceptibility gene even though the null hypothesis (no gene) is true. A false negative result occurs when the null hypothesis is accepted even though the gene is actually involved in the disorder. A further complication arises from the fact that a large number of genetic markers are being investigated, raising the problem of multiple testing. For linkage analysis it was assumed that 500 markers were genotyped, whereas for association studies it was assumed that 500,000 markers were studied. Given the typical distance over which linkage and association usually extend, it was further assumed that each of these tests was independent and a correction was made accordingly. In all the power calculations it was assumed that the genome-wide false positive rate was 5% and the false negative rate was 20% (after Risch and Merikangas [1, 9]).

RESULTS

The sample size necessary to achieve a given statistical power is dependent on both the magnitude of the genetic effect (λ_s), and on the frequency of the susceptibility allele (p). These two factors need not be independent. For example, a fully penetrant, causative allele will face much stronger selection pressure than an allele that only results in a modest increase in the probability of disease. As such, most likely it will have a lower frequency in the general population. Table 1 shows that the frequency of the susceptibility allele can dramatically affect the sample size. Because the allele frequencies will usually be unknown, accounting for this in the design of experiments will be difficult.

When the λ_s for a particular susceptibility locus is small, the number of individuals who must be typed for a classical sibling-pair genome scan is far greater than the number required for an association scan with the same power (Table 1). Thus, as was asserted by Risch and Merikangas [1], association studies appear to be more effective for detecting loci of moderate or small effect. However, if the λ_s for a given locus is higher (say >2), this is no longer always the case, with linkage studies usually requiring the collection and genotyping of fewer families [10]. For example, if a susceptibility gene has a λ_s of 1.5 and a population frequency of 1%, 338 sibling-pairs are required for a linkage screen to have an 80% probability of reaching the genome-wide 5% significance

Table 1. The necessary sample size to detect a significant result ($P < 0.05$) with 80% power for linkage analysis and TDT

γ	λ_S	p	Linkage (sibpairs)	TDT (trios)
8	1.5	0.01	338	375
	10.3	0.1	53	63
4	1.1	0.01	6400	1098
	2.0	0.1	276	150
2	1.01	0.01	445838	5823
	1.1	0.1	8085	695
1.5	1.002	0.01	6940000	19320
	1.02	0.1	101897	2218

Abbreviations are: TDT, transmission disequilibrium test; p, the frequency of the mutation; λ_S, the ratio of the recurrence risk in siblings to the population risk; γ, the genotype relative risk (after Risch & Merikangas, 1996, 1997 [1,9]).

level compared to 375 trios for a transmission disequilibrium test (TDT) of equal power. Given the λ_s for hypertension is approximately 3.5 [11], the relative merit of the two approaches depends on whether the genetic effect is due to a single major locus or spread across a number of different loci, possibly acting in an epistatic manner. For a given disease, if the alleles at one particular locus determine the majority of the genetic component of susceptibility, then a screen for linkage may be the best approach. However, if susceptibility is controlled by a number of loci, screening for association may provide better results. In insulin-dependent diabetes mellitus (IDDM), which has a λ_s of about 10, a locus near the human leukocyte antigen (HLA) region is a major determinate of susceptibility. This locus was easily found in a linkage screen using approximately 100 sibling-pairs [12]. Conversely, three whole-genome scans have been carried out in multiple sclerosis (which shows much greater familial aggregation with a λ_s of between 20 and 50). Each of these studies used approximately 100 sibling-pairs, yet none of them definitively localized a major susceptibility locus [13–15]. For this disorder, genome scans for linkage proved ineffective. In hypertension, a large number of genetic studies have been carried out for both linkage and association [reviewed in 16] with no single major locus being found [17]. This suggests that the genetic basis of the disease is controlled by a number of loci of modest effect, and hence genome-wide scans for linkage will probably be ineffectual. Whether a scan for association will prove more effective remains uncertain as such an experiment is impractical using current technology. When such technology does exist, the decision of whether to search for linkage or association will become an important issue.

ACKNOWLEDGMENTS

This paper was presented at the International Society of Nephrology 1997 Forefronts meeting in Germany. The author is supported by a Wellcome Trust International Prize Research Fellowship.

Reprint requests to Dr. Hywel B. Jones, MRC Biostatistics Unit, Cambridge University, Robinson Way, Cambridge, England CB2 2SR, United Kingdom. E-mail: hbj@mole.bio.cam.ac.uk

REFERENCES

1. Risch N, Merikangas K: The future of genetic studies of complex diseases. *Science* 273:1516–1517, 1996
2. Penrose LS: The detection of autosomal linkage in data which consists of pairs of brothers and sisters of unspecified parentage. *Ann Eugen* 6:133–138, 1935
3. Risch N: Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46:222–228, 1990
4. Risch N: Linkage strategies for genetically complex traits. II The power of affected relative pairs. *Am J Hum Genet* 46:229–241, 1990
5. Risch N: Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet* 46:242–253, 1990
6. Burke W, Motulsky AG: Hypertension, in *The Genetic Basis of Common Disease,* edited by King RA, Rotter JI, Motulsky AG, Oxford, OUP, 1992, pp 170–191
7. Hauser ER, Boehnke M, Guo SW, Risch N: Affected sibpair interval mapping and exclusion for complex genetic traits–Sampling considerations. *Genet Epi* 13:117–137, 1996
8. Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516, 1993
9. Risch N, Merikangas K: Genetic analysis of complex diseases–Response. *Science* 275:1329–1330, 1997
10. Scott WK, Pericak-Vance MA, Haines JL: Genetic analysis of complex traits. *Science* 275:384–385, 1997
11. Brown MJ: The causes of essential hypertension. *Brit J Clin Pharm* 42:21–27, 1996
12. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SCL, Jenkins SC, Palmer SM, Balfour KM, Rowe BR, Martin F, Barnett AH, Bain SC, Todd JA: A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371:130–136, 1994
13. Sawcer S, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J, Robertson N, Clayton D, Goodfellow PN, Compston A: A genome screen in multiple sclerosis reveals susceptibility loci on chromosomes 6p21 and 17q22. *Nature Genet* 13:464–468, 1996
14. Ebers GC, Kukay K, Bulman DE, Sadovnick AD, Rice G, Anderson C, Armstrong H, Cousin K, Bell RB, Hader W, Paty DW, Hashimoto S, Oger J, Duquette P, Warren S, Gray T, O'Conner P, Nath A, Auty A, Matz L, Francis G, Paulseth JE, Murray TJ, Pryse-Phillips W, Nelson R, Freedman M, Brunet D, Bouchard J-P, Hinds D, Risch N: A full genome search in multiple sclerosis. *Nature Genet* 13:4772–476, 1996
15. Haines JL, Ter-Minassian M, Bazyk A, Gusella JF, Kim DJ, Terwedow H, Pericak-Vance MA, Rimmler CS, Haynes CS, Roses AD, Lee A, Shaner B, Menold M, Seboun E, Fitoussi R-P, Gartioux C, Reyes C, Ribierre F, Gyapay G, Weissenbach J, Hauser SL, Goodkin DE, Lincoln R, Usuku K, Garcia-Merino A, Gatto N, Young S, Oksenberg JR: A complete genomic screen for multiple sclerosis underscores the role of the major histocompatibility complex. *Nature Genet* 13:469–471, 1996
16. Lifton RP: Molecular genetics of human blood pressure variation. *Science* 272:676–680, 1996
17. Jeunemaitre X, Soubrier F, Kotelevtsev YV, Lifton RP, Williams CS, Charru A, Hunt SC, Hopkins PN, Williams RR, Lalouel J-M, Corvol P: Molecular basis of human hypertension: Role of angiotensinogen. *Cell* 71:169–180, 1992